wwPDB NMR Structure Validation Summary Report (i)

Apr 20, 2024 – 08:14 AM EDT

PDB ID	:	5HPD
BMRB ID	:	30004
Title	:	Solution Structure of TAZ2-p53TAD
Authors	:	Krois, A.S.; Ferreon, J.C.; Martinez-Yamout, M.A.; Dyson, H.J.; Wright, P.E.
Deposited on	:	2016-01-20

This is a wwPDB NMR Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

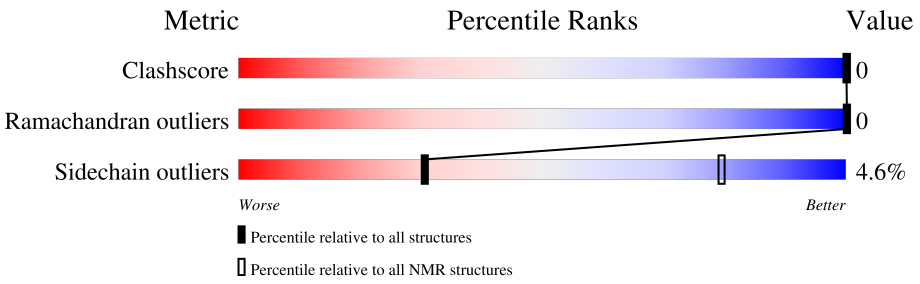
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 82%.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and a grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5%

Mol	Chain	Length	Quality of chain
1	A	158	63% 6% 31%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 8 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:1767-A:1853, A:3020-A:3026, A:3041-A:3055 (109)	0.35	8

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 2 clusters and 3 single-model clusters were found.

Cluster number	Models
1	1, 2, 5, 7, 8, 9, 11, 12, 14, 15, 16, 17, 18, 19, 20
2	4, 10
Single-model clusters	3; 6; 13

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2431 atoms, of which 1200 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called CREB-binding protein,Cellular tumor antigen p53 fusion protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	158	2428	753	1200	224	235	16	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2001	GLY	-	linker	UNP P45481
A	2002	SER	-	linker	UNP P45481
A	2003	GLY	-	linker	UNP P45481
A	2004	SER	-	linker	UNP P45481
A	2005	GLY	-	linker	UNP P45481
A	2006	SER	-	linker	UNP P45481

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	3	3	3

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

• Molecule 1: CREB-binding protein,Cellular tumor antigen p53 fusion protein

Chain A: 63% 6% 31%

S1764 P1765 Q1766 R1769 R1770 R1775 R1787 R1801 R1811 R1852 Q1853 Q1854 Q1855 G2001 S2002 G2003 S2004 G2005 S2006 E3002 E3003 P3004 Q3005 S3006 D3007 P3008 S3009 V3010 E3011 P3012 P3013 L3014 S3015 Q3016 E3017 T3018 F3019 W3023 P3027 E3028 N3029 N3030 V3031 L3032 S3033 P3034 L3035 P3036 S3037 Q3038 A3039 M3040

L3043 T3055 E3056 D3057 P3058 G3059 P3060 D3061

4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 8. Colouring as in section 4.1 above.

• Molecule 1: CREB-binding protein,Cellular tumor antigen p53 fusion protein

Chain A: 64% 5% 31%

S1764 P1765 Q1766 R1769 R1770 R1787 R1811 R1852 Q1853 Q1854 Q1855 G2001 S2002 G2003 S2004 G2005 S2006 E3002 E3003 P3004 Q3005 S3006 D3007 P3008 S3009 V3010 E3011 P3012 P3013 L3014 S3015 Q3016 E3017 T3018 F3019 W3023 P3027 E3028 N3029 N3030 V3031 L3032 S3033 P3034 L3035 P3036 S3037 Q3038 A3039 M3040 L3043 T3055

E3056 D3057 P3058 G3059 P3060 D3061

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing, molecular dynamics.*

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy.*

The following table shows the software used for structure solution, optimisation and refinement.

Software name	**Classification**	**Version**
CYANA	structure calculation	2.1
Amber	refinement	12

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	1735
Number of shifts mapped to atoms	1735
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	82%

6 Model quality (i)

6.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	A	0.62±0.00	0±0/888 (0.0± 0.0%)	1.06±0.02	5±1/1195 (0.5± 0.1%)
All	All	0.62	0/17760 (0.0%)	1.06	109/23900 (0.5%)

There are no bond-length outliers.

5 of 15 unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	1770	ARG	NE-CZ-NH2	7.60	124.10	120.30	6	17
1	A	1811	ARG	NE-CZ-NH2	7.41	124.01	120.30	16	6
1	A	1769	ARG	NE-CZ-NH2	7.25	123.93	120.30	4	18
1	A	1775	ARG	NE-CZ-NH2	6.89	123.74	120.30	5	11
1	A	1787	ARG	NE-CZ-NH2	6.78	123.69	120.30	11	18

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
All	All	17480	17680	17640	-

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including

hydrogen atoms). The all-atom clashscore for this structure is -.

There are no clashes.

6.3 Torsion angles (i)

6.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the backbone conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	109/158 (69%)	105±1 (96±1%)	4±1 (4±1%)	0±0 (0±0%)	100	100
All	All	2180/3160 (69%)	2099 (96%)	81 (4%)	0 (0%)	100	100

There are no Ramachandran outliers.

6.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the sidechain conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	102/146 (70%)	97±1 (95±1%)	5±1 (5±1%)	31	79
All	All	2040/2920 (70%)	1946 (95%)	94 (5%)	31	79

5 of 21 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	3023	TRP	20
1	A	3043	LEU	18
1	A	3055	THR	16
1	A	1837	ASN	10
1	A	1801	ARG	7

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 82% for the well-defined parts and 83% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *taz2p53_corrected_assignments.str*

7.1.1 Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	1735
Number of shifts mapped to atoms	1735
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	152	-0.26 ± 0.13	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	145	0.28 ± 0.07	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	130	-0.49 ± 0.26	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 82%, i.e. 1253 atoms were assigned a chemical shift out of a possible 1523. 0 out of 17 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	404/538 (75%)	205/216 (95%)	106/218 (49%)	93/104 (89%)
Sidechain	785/895 (88%)	535/577 (93%)	237/272 (87%)	13/46 (28%)

Continued on next page...

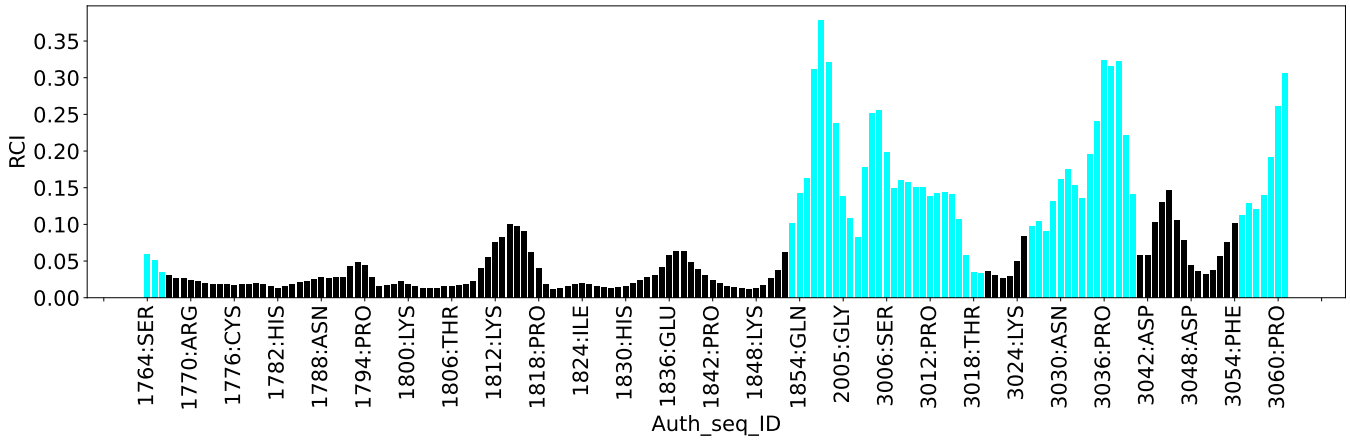
Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	64/90 (71%)	34/46 (74%)	28/35 (80%)	2/9 (22%)
Overall	1253/1523 (82%)	774/839 (92%)	371/525 (71%)	108/159 (68%)

7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: